Report on ANN for neutrino selection & Status report

- We present result from the construction of the ANN that "selects" neutrino interactions.

- We also present the results of the previous ANN on events from nustrip files of periods 1 and periods 2 that no neutrino interactions were found.

- We give a status report on the SF clustering and on the other ANN's for neutrino event classification

ANN for neutrino selection

- **Goal :** Construct an ANN that will be able to distinguish **neutrino interactions** from **not neutrino interactions**.

- **Procedure :**
  - **i)** Use the **existent neutrino interactions** as the **"signal"** training set and **interactions from nustrip files** as the **"background"** training set.
  - ii) **Redecode** and **reanalyze** (using Bruce's new vertex finding routine) neutrino and not neutrino events since we observed **changes** on the **SF and VDC decoders** that did not affect all events
  - iii) Construct **two ANNs** with the same "signal" training sets but different "background" sets to check the its performance and stability, since the number of training events is limited.

ANN Input Variables

- The input variables we have used to train the ANN are the same with the ones we are using for neutrino interaction identification plus 4 new (22 in total) :

  - TDC value differences T3-T2,T2-T1,T3-T1 **NEW**
  - Calorimeter energy along y=0 and $|x| > 100$ cm **NEW**
  - Number of SF, DC, VDC, MID hits,
  - Total Pulse height, % of SF hits in Stations 1 2 3 & 4
  - Total Energy in the EMCAL, number of Clusters, Average Cluster Energy,mean angle of clusters with respect to the z-axis from the interaction vertex
  - Number of SF lines, DC tracks, Final tracks

ANN 1 & 2 Training set and structure

- The **"signal" training set** for both ANNs is the same , and consists of **~ 850 events** that are characterized as **neutrino interactions .**
- The "**background" training set** for **ANN1** consists of ~850 events from nustrip files:

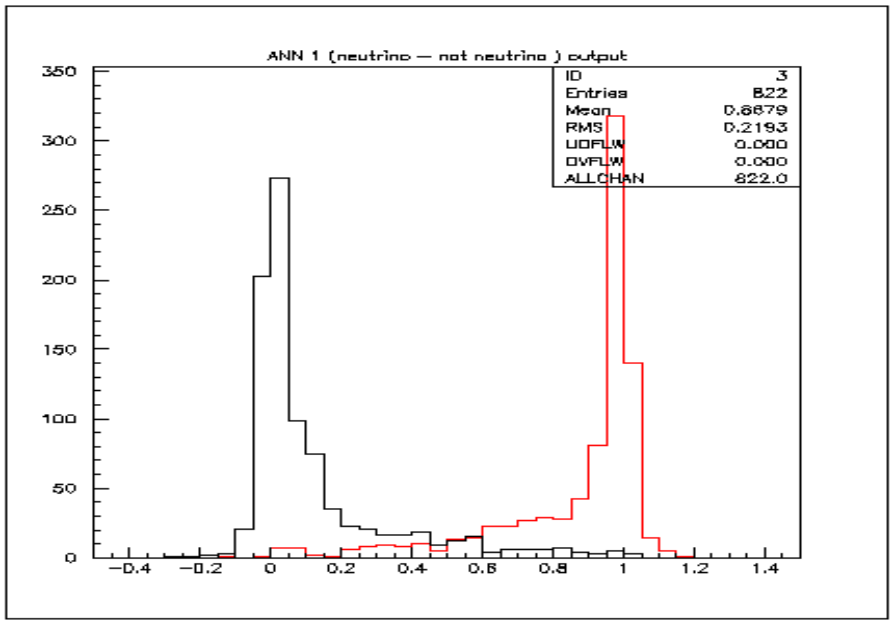
  2833 - 2927 - 2985 - 3143 - 3300 - 3331

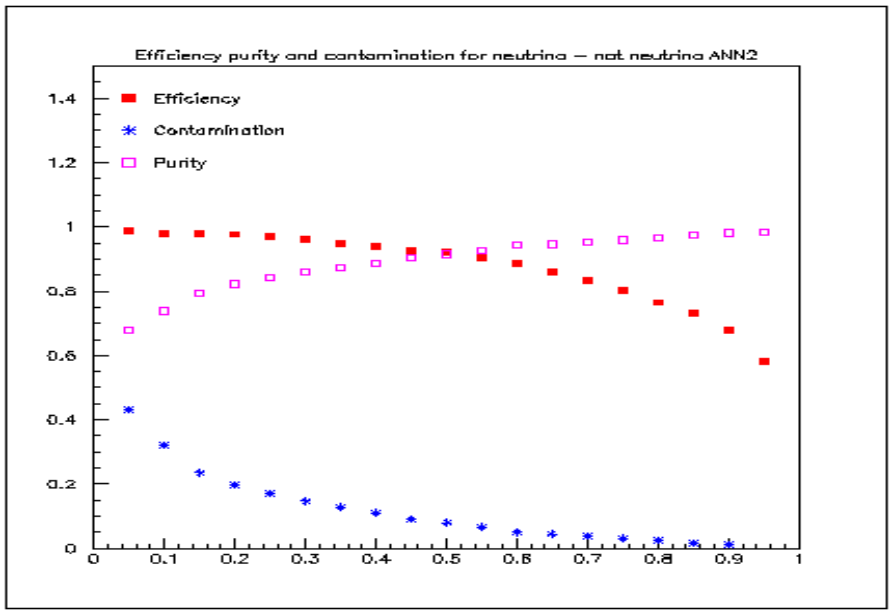
  and for **ANN2** consists of ~850 events from nustrip files:

  2911 - 3004 - 3137 - 3171 - 3222

- The network **structure** for both ANNs is **22 - 7 - 6 - 1**

Performance of ANN 1

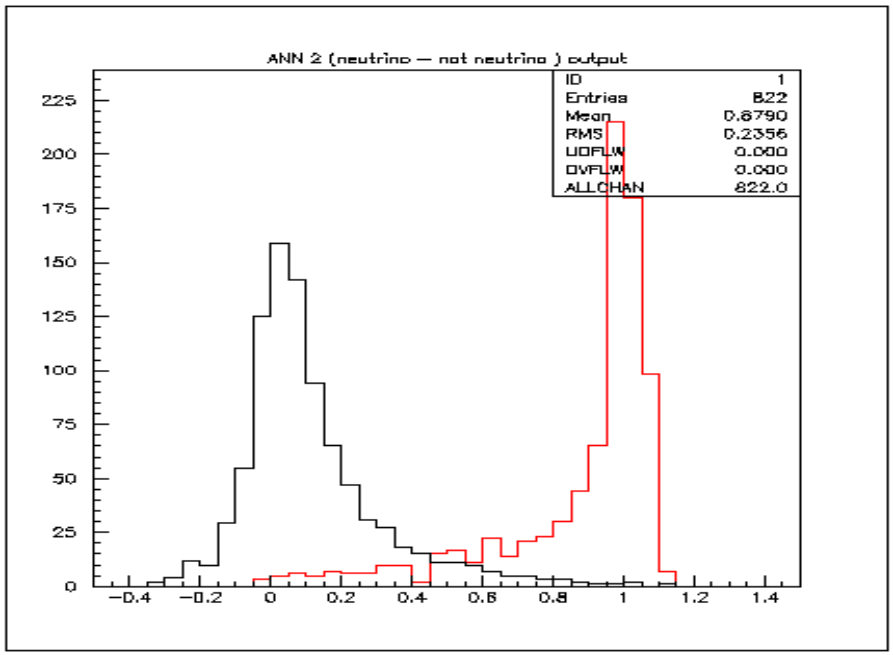


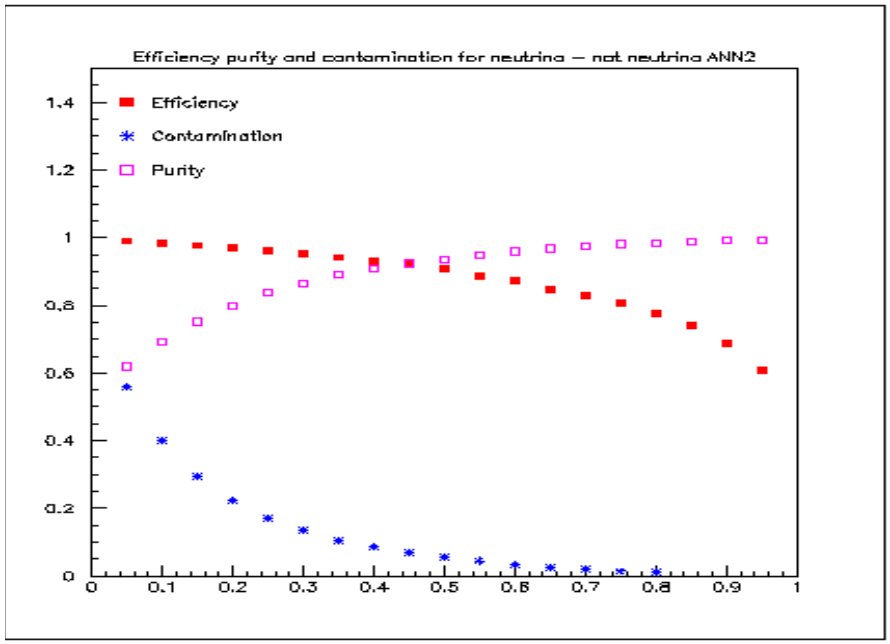


- The behavior of the ANN is very good and one select events with **high efficiency and high purity.**
- With a cut **@0.7** :

  Efficiency 83 % Purity 95 % Contamination 4 %

Performance of ANN 2





- The behavior of this second ANN is also very good and again we can select events with **high efficiency and high purity.**
- With a cut **@0.7** :

Efficiency 83 % Purity 97 % Contamination 2 %

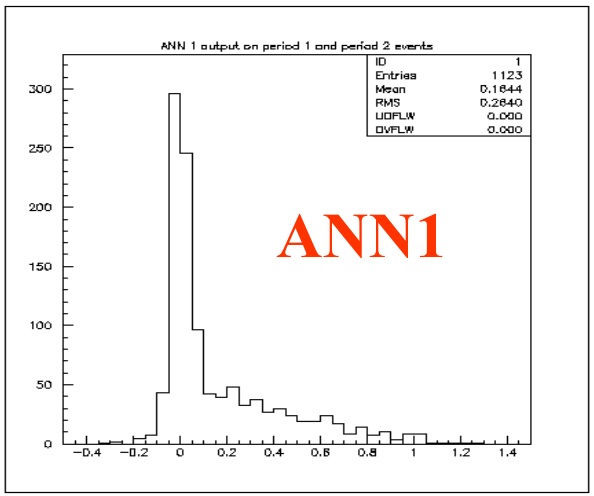
Significance of input variables for ANN1&2

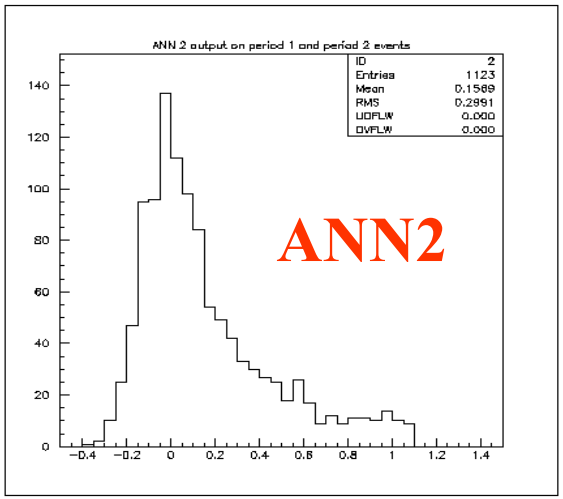
- **The significance** of each **input variable** is related with the **total weight** that this variable has when the network is trained.
- For ANN 1 & 2 the relative % weight for the **ten first variables** is :

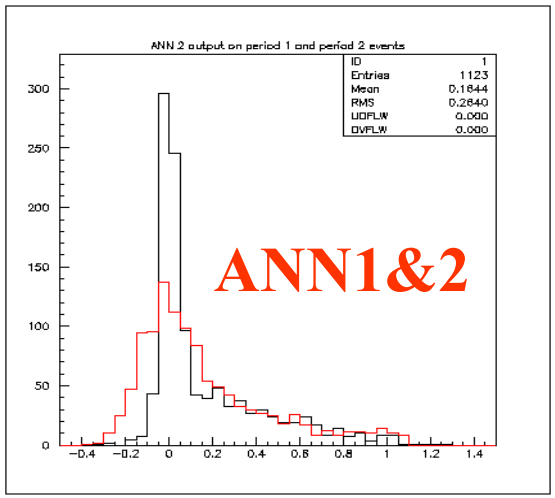
# SF lines	**weight 8.4647980**	**# of final tracks with hits in the SFs and DCs**	**weight 10.2959671**
# of final tracks with hits in the SFs and DCs	**weight 8.0314884**	**# EMCAL clusters**	**weight 8.6211538**
# of final tracks	**weight 7.2680130**	**# SF lines**	**weight 8.1007071**
Tot PH	**weight 7.2711334**	**# SF hits**	**weight 6.4359331**
# DC hits	**weight 6.4406667**	**# of final tracks**	**weight 6.3324022**
# EMCAL clusters	**weight 6.3714218**	**% hits in SF1**	**weight 6.1016245**
# SF hits	**weight 5.7427559**	**Tot. PH**	**weight 5.6125531**
T3 - T2	**weight 5.2126207**	Cluster angle	weight 5.0546441
% hits in SF1	**weight 5.1941943**	**# DC hits**	**weight 4.8059692**
# Center MID hits	weight 4.8600311	**T3 - T2**	**weight 4.1927462**

- We observe that the **10 most significant** variables are the **same for both ANNs**

Results of ANN1 & 2 for period1 and 2 events







- Mainly for checking the **performance and the stability** of the ANN we have used **period 1 and 2 events** from nustrip files in fmss that no neutrino interactions were found and **applied the ANNs selection functions**.
- With a **0.7 cut ANN1 selected 66** events and **ANN2 selected 86 events**.
- **64 % of the selected events** from the 2 ANNs are **identical** and that ratio goes **up to 74 %** if we consider as "selected" events with **probabilities 0.64 - 0.69.**

Conclusions

- Both **ANN1 and ANN2** show a very good **discriminate power** for **selecting neutrino events**.

- The fact that :

a) The **performance** (output functions) **is almost identical for both ANNs**

b) The **significance of the input variables** is **very similar**

c) The **results on period1 and period2 events** are **comparable.**

allows us to say that the **neutrino selection network** is **quite reliable**.

- We need to perform **a few more tests** using **different "background" sets** and also **examine** the **performance of the network** in **period3 and period4 events.**

Status report

- As far as **SF clustering** is concerned we are in the phase of **tuning our cuts with MC events** in order to select the final ones.

- We are going to use **all neutrino events** and **apply the selection functions** we have obtained from the **ANNs we have created so far** and see their results.

- For the improvement of the **ANNs for neutrino identification** we are using **Bruce's new MC code** to produce **simulated daft m-files** and we have started **looking at emulsion related variables**.

- In order to **use simulated emulsion information** we need to **reproduce MC files** with their associated **daft m files**, **include** the new **emulsion related variables** and **examine** the **performance** of the **existent ANNs** and also **construct new ones.**